

Glenn Roe (Australian National University)

L'étude littéraire à l'ère du numérique: du texte à l'intertexte dans les "digital humanities"

An integral part of the growing international field of concerns and practices in and around the "Digital Humanities", computational and digitally-assisted approaches to humanities text collections have in the past several years become increasingly prevalent. This article aims to explore both the promise and perils of these new digital approaches to literary and historical data sets, as well as the methodological underpinnings that inform their elaboration. In particular, we will examine several computational approaches to literary-historical analysis in French literature undertaken by the ARTFL Project at the University of Chicago, one of the oldest North American centres for digital humanities research.

S'il est vrai qu'un mouvement littéraire, ou bien savant, soit parvenu sur la scène dès qu'un "manifeste" a été produit, alors on peut dire que les "Digital Humanities", ou bien les "humanités numériques",¹ sont arrivées officiellement en France avec la publication (en ligne, bien sûr) du "Manifeste des Digital humanities" lors de la 'non-conférence' THATCamp qu'a eu lieu à Paris le 18–19 mai 2010 (Collectif 2011). Se distinguant d'un texte analogue publié en 2009 à l'Université de Californie-Los Angeles (Collectif s.d.) – d'une teneur beaucoup plus radical et contestataire, et ainsi moins pertinent hors de son contexte américain – le "Manifeste des Digital humanities" francophone s'organise dans sa présentation autour de trois pôles fondamentaux :

1. Le tournant numérique pris par la société modifie et interroge les conditions de production et de diffusion des savoirs.
2. Pour nous, les digital humanities concernent l'ensemble des Sciences humaines et sociales, des Arts et des Lettres. Les digital humanities ne font pas table rase du passé. Elles s'appuient, au contraire, sur l'ensemble des paradigmes, savoir-faire et connaissances propres à ces disciplines, tout en mobilisant les outils et les perspectives singulières du champ du numérique.
3. Les digital humanities désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales. (Collectif 2011)

¹ En suivant Aurélien Berra, je me sers du terme "humanités numériques" au lieu du double néologisme "digital humanities" transplanté tel quel en français. Pour une discussion plus soutenue des enjeux intellectuels des "humanités numériques" francophones, voir l'excellent carnet (blog) d'Aurélien Berra, *Philologie à venir* (Berra 2010–).

Soit. Mais, à part les conditions de production et de diffusion des savoirs, repensées pour s'engager mieux avec la culture montante du numérique, et la très nécessaire valorisation des approches traditionnelles des sciences humaines et sociales mobilisées au champ numérique – points essentiels sans doute dans la conception de cette 'transdiscipline' numérique et humaniste – quelles sont en fait ces méthodes, ces dispositifs, et ces perspectives heuristiques susmentionnés, liés pour ainsi dire, "au numérique dans le domaine des sciences humaines et sociales" ? En d'autres termes, quelles sont les implications pour la recherche littéraire, historique, culturelle ('humaniste', enfin) d'une vingtaine d'années d'efforts de numérisation – à la fois de la part des savants, des bibliothèques nationales et universitaires et des entreprises commerciales – censés remplir cette 'archive infinie' qu'est devenu l'internet et la bibliothèque numérique mondiale ? Quel sera enfin le rôle des méthodes et des algorithmes développés dans le domaine de l'informatique dans l'élaboration et l'interrogation des collections numériques d'une portée évidemment humaniste ? Questions pertinentes, auxquelles les 'humanités numériques' doivent essayer de répondre.

1 Introduction

Situées à la rencontre des méthodes informatiques avec des ressources humanistes et culturelles – entre l'homme et la machine, on pourrait dire, l'humain et le numérique – les 'humanités numériques' auxquelles je veux faire référence peuvent être conçues (dans un sens assez restreint, il faut l'admettre) comme l'application de nouvelles méthodes numériques à l'étude, et aussi à l'enseignement, de la littérature et de l'histoire. Combinaison heureuse, si j'ose dire, des approches informatiques aux sciences humaines et sociales qui a rajeuni en quelque sorte la recherche humaniste anglo-américaine ces dernières années. C'est de cette méthodologie particulière et de ses conséquences pour le devenir des études littéraires à l'ère du numérique que je voudrais parler dans la présente contribution. Méthodologie qui est (ou bien, qui devrait être) à la base de ce nouveau champ d'étude et qui peut, en quelque sorte, nous aider tous à mieux cerner et effectuer ce tournant numérique dans l'étude historique et littéraire.

Cette évolution particulière dans les méthodes de recherche, et aussi dans les pratiques savantes en général (et peut-être même dans la lecture tout court), est une manifestation, sans doute, d'une transformation culturelle plus large, devenu un lieu commun de nos jours, du passage collectif de la culture de l'imprimé et du livre – telle qu'elle a été décrite par Elizabeth Eisenstein il y a déjà une trentaine d'années (Eisenstein 1979) – à la culture numérique dans laquelle nous vivons évidemment aujourd'hui (voir Doueïhi 2008 et 2011). Personne, peut-être, n'a mieux compris les enjeux intellectuels et socio-culturels de cette transformation de l'imprimé au numérique que l'éminent historien du livre Roger Chartier, qui la décrit ainsi :

Le déroulement séquentiel du texte sur l'écran, la continuité qui lui est donnée, le fait que ses frontières ne sont plus aussi radicalement visibles que dans le livre qui enferme à l'intérieur de sa reliure ou de sa couverture le texte qu'il porte, la possibilité pour le lecteur de mêler, d'entrecroiser, d'assembler des textes qui sont inscrits dans la même mémoire électronique : tous ces traits indiquent que la révolution du texte électronique est une révolution des structures du support matériel de l'écrit comme des manières de lire. (Chartier 1997: 13)

Or, ce dépassement des contraintes physiques du codex que représentent les textes numériques nous mène peut-être à remettre en question une fois encore cette vision essentiellement 'humaniste' (dans toute l'ampleur de ce mot) de la littérature. Que reste-t-il, en fait, des résonances fines des mots et des textes littéraires – cet écoulement linguistique pour ainsi dire, du phénomène littéraire à travers l'histoire – dès qu'on les transforme en textes numériques, qu'on les mêle, les entrecroise, et les assemble avec des milliers et des milliers d'autres textes dans la prolifération croissante de l'archive infinie ? Comment garantir l'historicité des textes et la textualité de l'histoire – pour faire référence au programme des *New Historicists* américains (Vesser 1989) – là où le "déroulement séquentiel" du texte numérique ne fait aucune distinction textuelle entre ses données ? Comment, enfin, pourrions-nous concevoir un certain "humanisme numérique" – pour emprunter le terme de Milad Doueïhi – qui ferait valoir à la fois la puissance de calcul de l'ordinateur (pour ne dire rien de sa puissance socialisante) et de la spécificité des lettres 'humaines' traduites en données et déformées, en quelque sorte, comme objet d'étude ?

Ces questions, comme l'on peut deviner, se trouvent à la base de toute une gamme de réflexions sur les 'humanités numériques' (ou bien sur ses formes antérieures ou complémentaires de 'Humanities Computing' ou de 'New Media Studies') et leur relation

avec les disciplines traditionnelles des sciences humaines et sociales. Questions qui remontent au milieu du siècle dernier et à la naissance de la computation scientifique, c'est-à-dire des algorithmes (notamment promus par Alan Turing) qui forment la base de l'informatique. Ce tournant quantitatif dans les sciences naturelles et physiques (pures et dures) a ainsi contribué à la répartition, décrite par C.P. Snow dans son article célèbre, entre "deux cultures" – l'une scientifique, l'autre humaniste – dans la vie intellectuelle de l'Occident (Snow 1993). Opposition binaire (et parfois arbitraire) entre les sciences dites 'qualitatives' et celles supposées 'quantitatives', cette coupure culturelle a marqué toute une génération de savants et de chercheurs universitaires. De nos jours, c'est une semblable fracture 'quantitative' ou bien 'informatique' ("the digital divide" comme on dit en anglais) dans les sciences humaines et sociales que les humanités numériques sont censées combler – tâche lourde, certes, mais qui fait penser aux défis méthodologiques affrontés et surmontés par l'histoire quantitative des années 1960–70 menée par un François Furet ou un Henri-Jean Martin par exemple.

Loin de remplacer les 'humanités traditionnelles', les humanités numériques offrent la possibilité d'amplifier, d'augmenter, et de rajeunir les disciplines humanistes dont elles proviennent, tout en forgeant de nouveaux liens interdisciplinaires. Et pourtant dans mon domaine, à savoir les études littéraires, persiste une certaine défiance face aux nouvelles technologies – je pense par exemple à Stanley Fish, qui a récemment critiqué les humanités numériques dans le *New York Times* (Fish 2011, 2012a, 2012b). Cette réticence qui touche un grand nombre de chercheurs, d'enseignants et même d'étudiants, peut être affrontée, me semble-t-il, de deux manières : tout d'abord, en soulignant la continuité historique et méthodologique des études et de la critique textuelles (que ce soit numérique ou non-numérique) ; et ensuite, en insistant sur le fait que l'approche numérique ne saurait se substituer à une solide formation littéraire dans ses dimensions historiques, théoriques, et critiques. Cette formation est d'une importance capitale car elle permet de résister à une facilité superficielle des pratiques informatiques, c'est-à-dire à l'utilisation des outils numériques comme des fins en soi.

Ces réflexions sur la recherche informatique et les humanités numériques nous démontrent, une fois encore, cette tension fondamentale dans les sciences (humaines ou naturelles) entre la tradition et l'innovation ; tension dynamique qui est indissociable de l'élaboration récente des humanités numériques (tension même évoquée dans son oxymore nominal de 'digital humanities') ; tension, enfin, que j'aimerais explorer par le biais de l'interaction numérique du texte, du contexte, et de l'intertexte (on pourrait facilement y ajouter l'hypertexte, mais cela romprait la règle de trois).

Avant de passer à ces trois points, j'ajouterais que si je peux apporter quelque chose à la discussion sur les humanités numériques, c'est à partir d'une perspective unique, pour ainsi dire, formée non seulement dans la réflexion sur et l'interaction avec le numérique, mais aussi dans cette dialectique du développement technologique et de la recherche scientifique. Et que si je m'appuie beaucoup dans cet article sur les outils développés autour du moteur de recherche PhiloLogic du projet ARTFL, ce n'est pas pour faire du prosélytisme – il existe d'autres manières d'interroger les textes numériques, et de très bonnes. C'est simplement que mon expérience n'est pas là, et qu'ayant participé dans la conception, dans l'élaboration et dans le déploiement de PhiloLogic, j'ai été exposé pour ainsi dire à une vision plus compréhensive des avantages et des inconvénients de ses applications à l'étude littéraire.² Tout logiciel est déterminé par les préjugés de son développeur, et chaque décision entreprise dans l'élaboration d'un outil représente autant de fonctionnalités gagnées que d'éléments perdus. L'important, pour nous, c'est d'avoir un libre accès à ces préjugés, c'est d'être capable de les critiquer et de les améliorer. À la différence des entreprises comme Google, dont les algorithmes de recherche demeurent en quelque sorte des formules magiques dont on n'a pas la moindre idée, la transparence des outils et du code dans les humanités numériques nous donne l'occasion de participer à leur élaboration, afin qu'ils se conforment autant que possible aux exigences de la communauté scientifique. Les approches numériques au texte, au contexte, et à l'intertexte que je vais montrer par la suite sont les miennes, mais il y a une multitude d'autres. Et c'est préci-

² Sur le Projet ARTFL de l'Université de Chicago, voir [<http://artfl-project.uchicago.edu/>] ; sur le logiciel PhiloLogic, voir [<https://sites.google.com/site/philologic3/home>].

sément dans cette pluralité méthodologique que les humanités numériques trouvent leur force.

2 Texte

Pour revenir littéralement au texte, la première phase, disons, dans le devenir des études littéraires ou textuelles à l'ère numérique s'est focalisée naturellement sur le texte comme *locus operandi* et sur la nouvelle textualité numérique – ce "radiant textuality" dont parle Jérôme McGann (2004), par exemple. Phase qui héritait directement de la culture de l'imprimé et du livre, à cause sans doute d'un certain fétichisme de l'objet matériel du livre, qui voulait reprendre l'héritage et même le vocabulaire de l'imprimé en l'appliquant au numérique – pensons aux 'pages' web, aux 'bookmarks' ou signets des navigateurs, etc. C'est le moment aussi de l'essor des premières éditions numériques (très spécialisées en nature), et avec elles, les premiers systèmes de balisage – comme la TEI (*Text Encoding Initiative* [<http://www.tei-c.org>]) – censés faciliter la représentation virtuelle des textes imprimés sur l'écran.

Ces premiers efforts pour informatiser l'étude et la critique textuelle ne différaient pas, au moins d'une façon significative, des méthodes traditionnelles et avaient même une certaine teneur rétrograde, je dirais, qui faisait penser à une sorte de 'New Criticism informatique'. Et en fait, à bien des égards, la position critique de ces premiers savants du texte numérique représentait un retour à la rigueur philologique de la critique universitaire du XIX^{ème} et XX^{ème} siècles, reflétant une profonde fidélité à la lettre au moment même où l'ère du soupçon herméneutique avait lieu. Mais, avec l'avènement de l'internet, la textualité numérique s'affranchit et devient plus vivante, plus diffusée, moins fixe, et en quelque sorte, omniprésente. La portée des études littéraires par ordinateur, aussi bien que la complexité de ses outils et l'ampleur de ses collections, faisait face alors à cette nouvelle réalité numérique.

Les nouvelles méthodes de recherche réalisées grâce aux premiers efforts des praticiens de ce que l'on appelait à l'époque 'Humanities Computing', et de la nouvelle textualité promue par l'internet, peuvent être considérées comme une alliance de la spécificité des mots (ou bien de la littérature) à travers l'histoire et de cette maniabilité technique du texte numérique dont parle Roger Chartier. Pour en donner un

exemple, je présenterai une problématique à laquelle je m'intéresse actuellement et que j'ai explorée avec l'aide de PhiloLogic. Mais, on pourrait facilement faire les mêmes sortes d'interrogation avec les outils *Voyant* [http://voyant-tools.org] développés par Stéfán Sinclair et Geoffrey Rockwell, ou bien avec d'autres moteurs de recherche informatiques. Disons, par exemple, que si vous êtes intéressés, comme je le suis, aux acceptions différentes du mot 'moderne' (c'est à dire, tout ce qui est moderne, modernisme, modernité, etc.) chez un écrivain 'anti-moderne' tel que Charles Péguy – en empruntant la désignation d'Antoine Compagnon (2005) – PhiloLogic vous fournit facilement, dans son contexte, toutes les occurrences du lemme 'modern*' tirés des œuvres de Péguy (fig. 1) :³

```
VMDug* (bibp.832)es bleu gris; ces durs aciers modernes; ces jantes grâles; ces aciers gris bleu; ces
Argen* (bibp.1102)l'âge chrétien. Il y a l'âge moderne. Une ferme en Beaune, encore après la guerre,
VMDug* (bibp.839) la France, même dans son âge moderne, ne laissera peut-être pas perdre cette gloire
VMDug* (bibp.799)une tragédie une action aussi moderne que celle-ci, si elle s'étoit passée dans le
VMDug* (bibp.831)petits jeunes gens de canons modernes, ces gringalets de canons modernes au corps d'
VMDug* (bibp.831)nes, ces gringalets de canons modernes au corps d'insecte, aux rouses comme des pattes
VMDug* (bibp.831)mons de 75, nos grâles canons modernes, si périlleux, ne peu trop lourds toutefois p
VMDug* (bibp.750)de l'une de nos chemises modernes, il la voyait ce qu'elle est: la pierre du foy
VMDug* (bibp.744)nie païen). Il avait le cœur moderne, ce qui est une deuxième façon de dire qu'il
VMDug* (bibp.799)ses millions, ses impôts déjà modernes sont très sensiblement inférieurs à Racine
Argen* (bibp.1123)iques. Et des anciens et des modernes. Pourvu qu'un homme parle de la sâtière classe
Argen* (bibp.1186)uveau, plein d'enjolivements modern style, au bas de la grande page du samedi de la
Argen* (bibp.1133)oisément le contraire d'être moderniste et c'est par un incroyable abus de langage q
VMDug* (bibp.646)lité antique et la fraternité moderne. Je ne serais point seulement un ingrat. Je n'a
Argen* (bibp.1122)nes modernes et généralement modernistes. Ils ne sont aucunement, et encore moins qu
VMDug* (bibp.711)aphie ancienne et une graphie moderne, je ne pense point qu'elles aillent jusqu'à ce
Argen* (bibp.1114)l'able mécanisme de la grève moderne à jet continu, qui fait toujours monter les sa
Argen* (bibp.1223) aux principes de la guerre moderne. Il n'y avait que les radicaux, dans ce temps-l
Argen* (bibp.1223) aux principes de la guerre moderne. Vous êtes trop jeunes, mes enfants, pour avoi
Argen* (bibp.1122) essentiellement des hommes modernes et généralement modernistes. Ils ne sont auc
Argen* (bibp.1170)Monsieur Rodier, l'invention moderne, la nouveauté moderne, ce n'est point l'exacti
VMDug* (bibp.713) l'écrire tranquillement à la moderne, sans aucune affectation. C'est ainsi que je tr
Argen* (bibp.1134)et un système de courage. Le modernisme est la vertu des gens du monde. La liberté
Argen* (bibp.1134) que l'adversaire croit. Le modernisme est un système de complaisance. La liberté
Argen* (bibp.1134)ai croie ce qu'il croit. Le modernisme consiste à ne pas croire soi-même pour ne
VMDug* (bibp.711)ogère, professionnel, dans le moderne), une certaine supériorité, une sorte de myti
Argen* (bibp.1134) un système de déférence. Le modernisme est un système de politesse. La liberté es
Argen* (bibp.1134) grande mots, mais enfin le modernisme est un système de lâcheté. La liberté es
Argen* (bibp.1134) mots. Le modernisme est, le modernisme consiste à ne pas croire ce que l'on croit.
Argen* (bibp.1134) grâce). Disons les mots. Le modernisme est, le modernisme consiste à ne pas croire
Argen* (bibp.1134) je ne hais rien tant que le modernisme. Et je n'aime rien tant que la liberté. (Et
```

Fig. 1 : Les occurrences du lemme 'modern*' tirés des œuvres de Charles Péguy

Résultats que vous pouvez aussi explorer dans un contexte plus large, ici quelques occurrences du mot 'modernisme' (fig. 2) :

```
39. Péguy, Charles. L'argent [Page 1134 | paragraphe | Section]
sort toujours ceux qui réussissent le mieux. Et c'est là une incroyable confusion. Et je ne hais rien tant que le modernisme. Et je n'aime rien tant que la liberté. (Et en elle-même, et n'est-elle point la condition irrévocable de la grâce). Disons les mots. Le modernisme est, le modernisme consiste à ne pas croire ce que l'on croit. La liberté consiste à croire ce que l'on croit et à admettre, (au fond, à exiger), que le voisin aussi croie ce qu'il croit. Le modernisme consiste à ne pas croire soi-même pour ne pas léser l'adversaire

40. Péguy, Charles. L'argent [Page 1134 | paragraphe | Section]
qui réussissent le mieux. Et c'est là une incroyable confusion. Et je ne hais rien tant que le modernisme. Et je n'aime rien tant que la liberté. (Et en elle-même, et n'est-elle point la condition irrévocable de la grâce). Disons les mots. Le modernisme est, le modernisme consiste à ne pas croire ce que l'on croit. La liberté consiste à croire ce que l'on croit et à admettre, (au fond, à exiger), que le voisin aussi croie ce qu'il croit. Le modernisme consiste à ne pas croire soi-même pour ne pas léser l'adversaire qui ne croit pas

41. Péguy, Charles. L'argent [Page 1134 | paragraphe | Section]
la condition irrévocable de la grâce). Disons les mots. Le modernisme est, le modernisme consiste à ne pas croire ce que l'on croit. La liberté consiste à croire ce que l'on croit et à admettre, (au fond, à exiger), que le voisin aussi croie ce qu'il croit. Le modernisme consiste à ne pas croire soi-même pour ne pas léser l'adversaire qui ne croit pas non plus. C'est un système de déclinisme mutuel. La liberté consiste à croire. Et à admettre, et à croire que l'adversaire croit. Le modernisme est un système de complaisance.

42. Péguy, Charles. L'argent [Page 1134 | paragraphe | Section]
aussi croie ce qu'il croit. Le modernisme consiste à ne pas croire soi-même pour ne pas léser l'adversaire qui ne croit pas non plus. C'est un système de déclinisme mutuel. La liberté consiste à croire. Et à admettre, et à croire que l'adversaire croit. Le modernisme est un système de complaisance. La liberté est un système de respect. Il ne faudrait pas dire les grands mots, mais enfin le modernisme est un système de lâcheté. La liberté est un système de courage. Le modernisme est la vertu des gens du monde. La liberté est

43. Péguy, Charles. L'argent [Page 1134 | paragraphe | Section]
pas léser l'adversaire qui ne croit pas non plus. C'est un système de déclinisme mutuel. La liberté consiste à croire. Et à admettre, et à croire que l'adversaire croit. Le modernisme est un système de complaisance. La liberté est un système de respect. Il ne faudrait pas dire les grands mots, mais enfin le modernisme est un système de lâcheté. La liberté est un système de courage. Le modernisme est la vertu des gens du monde. La liberté est
```

Fig. 2 : Les occurrences du mot 'modernisme' chez Péguy

³ Tous les résultats qui suivent sont tirés de la base de données ARTFL-Frantext [http://artfl-project.uchicago.edu/content/artfl-frantext] du projet ARTFL de l'Université de Chicago.

Une navigation qui est, comme on peut voir, enracinée dans la terminologie de l'imprimé avec des liens aux 'pages' aux 'paragraphes' aux 'chapitres', etc. et qui vous permet enfin de tomber sur la page même dans *L'Argent suite* où Péguy, bon anti-moderne qu'il est, constate que "le modernisme consiste à ne pas croire ce que l'on croit", phrase typiquement péguyste d'ailleurs.

À côté de cette fidélité à l'œuvre et à sa cohérence textuelle – si nécessaire d'ailleurs à l'étude littéraire – l'une des commodités de la recherche numérique, c'est l'oscillation entre cette forme de lecture précise et immersive d'un mot et de son contexte, et la considération plus large des champs lexicaux dans lesquels ces mots opèrent. Ici, par exemple, le logiciel génère un tableau de collocation du mot 'moderne' dans toutes ses formes chez Péguy (fig. 3) :

	Within 5 Words on Either Side	Within 5 Words to Left only	Within 5 Words to Right only
1	temps (225)	temps (211)	partie (33)
2	monde (138)	monde (122)	celle (22)
3	science (127)	science (112)	ceux (21)
4	société (99)	société (93)	ancienne (21)
5	esprit (98)	société (80)	révolution (20)
6	histoire (90)	esprit (79)	notre (20)
7	société (87)	histoire (78)	chez (20)
8	notre (86)	homme (85)	esprit (19)
9	homme (85)	notre (84)	beaucoup (19)
10	chez (72)	philosophie (83)	— (18)
11	ancienne (71)	civilisation (82)	trop (18)
12	philosophie (70)	industrie (55)	seulement (18)
13	civilisation (66)	chez (52)	système (17)
14	industrie (60)	ancienne (50)	doit (17)
15	état (57)	idées (49)	celui (17)
16	économie (55)	état (48)	— (16)
17	idées (54)	économie (45)	homme (16)
18	nations (49)	nations (45)	France (16)

ailleurs autour ancien ancienne anciennes anciens
 anglais antique antique antique aujourd'hui autres auteurs
 avant avec beaucoup beaucoup beaucoup celle celle
 celui celui ces ceux chaque chez chimie chimie
 christianisme civilisation classique combat comment commune
 conditions conscience contemporaine contraire contre critique
 depuis deux doit deux droit découvertes déjà
 démocratie démocrates développement elle elle elle
 enseignement esprit état état eu Europe exemple six
 femme lui forme formes France français grand grande
 grandes grands grand guerre histoire homme
 hommes lui idée idées industrie introduction
 immortel jeune par jusqu'à langue langues dans littérature
 littérature livre livre livres mais modern moderne
 monde morale mot nous mouvement moyen
 nations nature notre oeuvre ou origine ou nombre parti
 parmi part partie pays pasteur peinture pensée pays
 peuples philosophes philosophie physique plus
 politique pourque pourque premier première principe
 principes production progrès propriété psychologie ou puis
 quelques raison nous nous roman régime révolution
 révolution science sciences sens sens sera servir
 seulement siècle socialisme société sociétés soit
 sur style surtout système tant tant temps temps
 thèse travail vers troisième trop trop type ville ville voir voir
 vraiment via âge être avec économie économie ligne
 époque étaient état état suite évolution —

Fig. 3 : Tableau de collocation du mot 'moderne*' dans toutes ses formes chez Péguy

Ce qui donne, peut-être, une idée de son usage et du champ lexical dans lequel il se trouve le plus fréquemment. On peut aussi comparer un tableau à un autre, par exemple la collocation des mots autour du lemme 'moderne' dans le grand roman de Marcel Proust, contemporain exact de Péguy (fig. 4) :

Keywords found (with occurrences): modern (5), moderne (27), modernes (18), modernisent (1), modernisme (1), moderniste (1), modernistes (2), modernité (1)
 The 118 most common words are being filtered from this report. To include filtered words select "Turn Filter Off" on the search-form.

	Within 5 Words on Either Side	Within 5 Words to Left only	Within 5 Words to Right only
1	style (6)	esprit (3)	style (5)
2	littérature (3)	paris (2)	littérature (2)
3	langage (3)	luxe (2)	langage (2)
4	esprit (2)	croyait (2)	évide (1)
5	vieux (2)	certain (2)	étrange (1)
6	presque (2)	étaient (1)	émervellants (1)
7	peuples (2)	équivalent (1)	égéries (1)
8	paris (2)	yacht (1)	voix (1)
9	moblier (2)	xviii (1)	vieux (1)
10	luxe (2)	voulait (1)	vers (1)
11	hôtel (2)	voir (1)	usure (1)
12	croyait (2)	vitruve (1)	truqué (1)
13	certain (2)	vieux (1)	tonifié (1)
14	beaucoup (2)	traditionnelle (1)	taillés (1)

absorbant affectation agi albertine **ancienne** anciennes années
 antique appartait apportait artillerie artiste artistes assez aucun
 esprit esprit sans esprit bagages bande baste basteux beau
beaucoup bégayé boucher bras carrière café certaines
certain aux chaude cherche chez clés coffre colonnes
 canby comment composé contemporain contenté contre couloirs
croyait demandes dicton directes d'ailleurs données donner
 drainage dressant due duchesse déclaration désiré désolé effier
 effluie entonnait entrât entré **esprit** au expositions
 exotiquement façon femme figurent file fort français gerant ghera
 genouille pile grille garnissantes guerres guerriers glire hauteur
hôtel hâblerie idées impaisible imposante infaisissables inactes
 intelligentes intelligentes intéressantes inutilisables inventions jusque
langage langues labourg labéris lesquelles libre
littérature l'univers **luxe** lyrique manifestations mangera
 mer meubles milieu nous **moblier** montent motif munituels
 nous nombreuses nupties notre séjour sources offertes saillir
paris parisiens part partie partien patois personnes **peuples**
 phalange phrases plénitude plier pleine portée pourvus possible
 pouvoir pour **presque** protestaire précipitamment
 préparations présente public puisent puzzle quiconque quelque
 régulariser rarissime reporter reproduire resta revêtir rivaliser roses
 répétait révisé rivaliser rivaliser s'échappait sainte senties sensor seulement
 sommes sources statues **style** stylisant superbes supporter
 surtout tableaux tapisseries toiles tonifié traditionnelle truqué usure
 vers **vieux** vitruve voir voix voulu xviii yacht égéries émervellants
 équivalent étaient étrange évite

Fig. 4 : Tableau de collocation du mot 'modern*' chez Marcel Proust

Pris ensemble (fig. 5), on notera peut-être (ce qui n'est pas étonnant étant donné nos deux auteurs) que pour Péguy, le 'moderne' est bel et bien un régime d'historicité, pour emprunter un terme à François Hartog (2003), un système temporel lié au charnel (le monde et les hommes) de l'existence, alors que pour Proust, on peut dire, qu'il y a un glissement vers le 'modernisme' plutôt littéraire, orienté vers la stylistique ou l'esthétique.

accorderais schwaer **actiers** action admirable adversaire affectation stance
 alors ancienne **anciens** au antique avouement agent art attendu
 assure automatisé avant aversures bas basteux **Canons** celle
 celui cersive **cet** aux chemins chez chrétien ci cinquante
 cœur commandements **commencement** complaisance comités
consiste constamment caribus contraires contraires **contre** corps
 couple courage crimes croire croit demande deuxième ils
 dans drapement duplicité dure différence **déjà** effectuer effet
 effet enjolivements **ensemble** entendre entendu au euripide aux
 auxseine forme format **formes** franco français fuyait inamitié fil
 gageure grov **gens** grands graphie gringalets gr gro grille
 grive grâles guerre généralement hâs hussards **homme**
 hommes hospitalité lui hâs hâs impéride imposeur inauguration
 inférieures ingratitude inexacte invention janses je jeunes joue
 jouer l'échéé malices maters **moderne** modernes
 modernisme modernistes **monde** monstrieux
 monstruosité mot **mots** mûricanisme mûlitude même nomment
 nous nouveau nouveauté nouveauté **paris** part parté paucres plein
 peine partie personnes petite peuple physiologique plein polémique
 politique population pouvaises **premièrement** principes
 professionnel pris **précisément** premier qu'onque regard nuller
 règne régner résuéré résuérance saints sauss sens senties sensé
 seul seulement sinon stusé société solent soit style sus surplombait
système tant **un** temporellement **temps**
 toulé tragédie trop turs **Vallée** vallent vertu valent vice **vivait**
 volé vu **âge** économique active également être

Péguy

absorbant affectation agi albertine **ancienne** anciennes années
 antique appartait apportait artillerie artiste artistes assez aucun
 esprit esprit sans esprit bagages bande baste basteux beau
beaucoup bégayé boucher bras carrière café certaines
certain aux chaude cherche chez clés coffre colonnes
 canby comment composé contemporain contenté contre couloirs
croyait demandes dicton directes d'ailleurs données donner
 drainage dressant due duchesse déclaration désiré désolé effier
 effluie entonnait entrât entré **esprit** au expositions
 exotiquement façon femme figurent file fort français gerant ghera
 genouille pile grille garnissantes guerres guerriers glire hauteur
hôtel hâblerie idées impaisible imposante infaisissables inactes
 intelligentes intelligentes intéressantes inutilisables inventions jusque
langage langues labourg labéris lesquelles libre
littérature l'univers **luxe** lyrique manifestations mangera
 mer meubles milieu nous **moblier** montent motif munituels
 nous nombreuses nupties notre séjour sources offertes saillir
paris parisiens part partie partien patois personnes **peuples**
 phalange phrases plénitude plier pleine portée pourvus possible
 pouvoir pour **presque** protestaire précipitamment
 préparations présente public puisent puzzle quiconque quelque
 régulariser rarissime reporter reproduire resta revêtir rivaliser roses
 répétait révisé rivaliser rivaliser s'échappait sainte senties sensor seulement
 sommes sources statues **style** stylisant superbes supporter
 surtout tableaux tapisseries toiles tonifié traditionnelle truqué usure
 vers **vieux** vitruve voir voix voulu xviii yacht égéries émervellants
 équivalent étaient étrange évite

Proust

Fig. 5 : Tableaux de collocation du mot 'modern*' chez Péguy et Proust

Tout cela pour dire qu'il ne s'agit pas, en ce qui concerne ces approches numériques de la littérature, d'un réductionnisme quelconque lié aux aspects purement quantitatifs des textes. Ni non plus d'une banalisation de la spécificité de l'œuvre littéraire – au contraire, ces approches facilitent la recherche littéraire, elles nous aident à tester des hypothèses, à confirmer (ou bien à réfuter) les intuitions intellectuelles, et à découvrir

des motifs, des rapports, et des allusions qui nous échapperaient peut-être sans la médiation numérique.

Ceci dit, il est néanmoins aisé de reconnaître les limites des recherches textuelles basés sur les mots-clés – et là on voit très bien que l'innovation d'hier devient vite la tradition d'aujourd'hui, comme disait d'ailleurs Stendhal des classiques et des romantiques. Si nous continuons notre investigation sur l'usage du 'moderne', en élargissant notre champ de recherche à tous les ouvrages dans la base de données Frantext venant de cette fin-de-siècle française (disons de 1880 à 1915), la tâche interprétative devient plus lourde :

	Within 5 Words on Either Side	Within 5 Words to Left only	Within 5 Words to Right only
1	temps (225)	temps (217)	partie (33)
2	monde (138)	monde (123)	celle (22)
3	science (127)	science (112)	ceux (21)
4	société (95)	société (75)	ancienne (21)
5	esprit (96)	société (80)	révolution (20)
6	histoire (90)	esprit (79)	notre (20)
7	société (87)	histoire (78)	chez (20)
8	notre (86)	homme (89)	esprit (19)
9	homme (85)	notre (84)	beaucoup (19)
10	chez (72)	philosophie (63)	— (18)
11	ancienne (71)	civilisation (62)	trop (18)
12	philosophie (70)	industrie (55)	seulement (18)
13	civilisation (66)	chez (52)	système (17)
14	industrie (60)	ancienne (50)	doit (17)
15	état (57)	idées (49)	celui (17)
16	économie (55)	état (48)	— (16)
17	idées (54)	économie (45)	homme (16)
18	nations (49)	nations (45)	france (16)

ailleurs amour ancien ancienne anciennes anciens
 angles antique antiques antique ART aujourd' aura auteurs
 avant nous beaucoup caractère catholique celle celui
 celui certain cet ceux chaque chez chimie chose
 christianisme civilisation classique combien comment commerce
 conditions conscience contemporaine contraire contre critique
 depuis dieu doit donc deux droit découvertes déjà
 démocratie démocratie développement elle elle enfin
 enseignement esprit avec eux au europe exemple six
 femme lui forme formes france français gens grande
 grandes grands guerre histoire homme
 hommes lui idée idées industrie introduction
 inventer justice jour jeter langues livres libéralisme
 littérature livre livre fait maître mais moderne
 monde morale mot nous mouvement moyen
 nations nature notre oeuvre or origine ou quartier Paris
 parmi pas partie pays peintre peinture pensée peut
 peuples philosophie philosophie physique plus
 politique pourquoi pourquoi puis premier première presque
 principes production progrès propriété psychologie au plus
 quelques raison raison revue roman régime révolution
 révolution science sciences sens sens sens sens
 seulement siècle socialisme société sociétés soit
 sorte style surtout système tant va temps
 théorie travail trois troisième trop troupe type y elle voir voir
 vraiment voir âge âge aux économie économie ague
 époque était état états étude évolution —

Fig. 6 : Tableau de collocation des formes de 'modern*' dans ARTFL-Frantext (1880–1915)

Ici (fig. 6), un tableau de collocation pour les presque 4 000 occurrences des formes de 'moderne' à travers plus de 500 documents. Et puis, venant plus loin, en considérant tout le XIX^{ème} siècle de 1815 à 1915, on est confronté avec presque 9 000 occurrences dans 1 500 documents. Et là, évidemment, le champ lexical devient 'un peu de tout dans tout' et la spécificité du mot-clé et de son contexte particulier se perdent progressivement de vue. En introduisant un nombre croissant de textes supplémentaires, cette dialectique entre le dire et le vouloir-dire, entre les textes et les contextes, commence à se défaire, et l'interprétation des résultats devient de plus en plus difficile. Donc, la lecture précise et attentive de ces 9 000 occurrences n'est pas tenable comme moyen sérieux de recherche littéraire. Et par là, alors, il devient évident qu'avec l'accroissement exponentiel des collections numériques, de nouvelles formes de lecture et de recherche informatique sont requises.

3 Contexte

Deuxième phase dans l'évolution de la recherche littéraire à l'ère du numérique, c'est celle qui résulte de la surcharge d'information et des changements correspondants qu'elle apporte à la textualité numérique. Nous venons de voir que la consultation des bases de données littéraires, même les plus soignées comme *ARTFL-Frantext* avec ses 3 500 livres, arrive vite aux limites fonctionnelles de la recherche informatique. Et, avec la prolifération progressive des ressources numériques, prise en charge, au début, par les institutions publiques et les bibliothèques universitaires, l'étendue des archives numériques globales dépasse, en quelques sortes, ses bornes institutionnelles. Après donc l'avènement de la textualité numérique, héritière fidèle de l'imprimé, vient de nos jours une contextualité croissante des textes venant de partout, non seulement des artefacts des livres imprimés mais aussi des textes 'nés numériques' (*born-digital*), des blogues, des articles en ligne, des courriels, etc.

Or, une des réponses possibles à cette surcharge de textes numériques et de leur contextes, réponse provocante sans doute, a été celle lancée par l'historien du roman Franco Moretti de Stanford University : c'est, en fait, *de ne plus lire* ces textes, ou plus précisément, *de les lire à distance* (*distant reading* en anglais). Comme son nom l'indique, cette lecture à distance s'oppose consciemment à la lecture continue et intensive (*close reading*) des critiques traditionnels :

Distant reading: where distance, let me repeat it, is a *condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, less is more. If we want to understand the system in its entirety, we must accept losing something. We always pay a price for theoretical knowledge: reality is infinitely rich; concepts are abstract, are poor. But it's precisely this 'poverty' that makes it possible to handle them, and therefore to know. This is why less is actually more. (Moretti 2000: 57-58)

La disparition du texte, évoquée par Moretti, est une réponse à la croissance du contexte numérique, dans lequel les œuvres individuelles se perdent sous la masse des autres textes et deviennent de moins en moins visibles. Pour Moretti et ses adeptes, la perte de spécificité de la chose littéraire devient l'occasion d'une nouvelle méthode d'étude, de cette 'lecture à distance' (à mon sens, 'non-lecture') fièrement quantitative de portée et plus concernée avec les métadonnées, c'est-à-dire avec les descriptions

formelles et informationnelles des textes qu'avec les données elles-mêmes et le texte qu'elles contiennent.

À ces nouvelles approches quantitatives de la littérature de masse (Moretti analysant les titres de tous les romans anglais publiés tout au long du XIX^{ème} siècle, par exemple) s'ajoute pendant les cinq dernières années et par l'entrée en scène des efforts de numérisation massive – gérés cette fois-ci par des entreprises commerciales tel que Google Books par exemple – une véritable explosion de données textuelles. Éclatement massif de livres numériques qu'a provoqué la question pertinente de Gregory Crane – directeur du projet Perseus, donc de la bibliothèque numérique des littératures grecques et latines : "Que faire avec un million de livres ?"

There are only about 30,000 days in a human life – at a book a day, it would take 30 lifetimes to read a million books and our research libraries contain more than ten times that number. Only machines can read through the 400,000 books already publicly available for free download from the Open Content Alliance. (Crane 2006)

Selon Crane, seules les machines peuvent lire ces millions de livres, ce qui est sans doute vrai, mais ce qui est peut-être aussi une autre manière de lire à distance, ou encore de ne pas lire. Question posée en 2006, quand les corpus numériques n'avaient pas encore atteint au millionième livre, alors que par la suite nous avons vu l'établissement d'un corpus de plus de 5 millions de livres (chiffres qui augmentent tous les jours). Résultat direct du projet Google Books, cette base textuelle énorme exige donc une autre manière de lire à distance, ou bien à l'ordinateur, une nouvelle 'discipline' (entre guillemets) de recherche informatique et linguistique qui s'appelle "culturomics". Effectué par une équipe de chercheurs à Harvard et avec la participation de plusieurs ingénieurs de Google, le travail informatique qui a engendré la "culturomique" a été le sujet d'un article très médiatisé dans la prestigieuse revue scientifique *Science*, avec des articles complémentaires dans le *New York Times*, la revue *Nature*, etc. :

We constructed a corpus of digitized texts containing about 4% of all books ever printed. Analysis of this corpus enables us to investigate cultural trends quantitatively. We survey the vast terrain of "culturomics" focusing on linguistic and cultural phenomena that were reflected in the English language between 1800 and 2000. We show how this approach can provide insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology. "Culturomics"

extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities. (Michel et al. 2010: 176)

Et là, nous voyons que cette méthode de recherche, marquée par une forte naïveté par rapport à la complexité linguistique et culturelle des textes, s'appuie exclusivement sur le quantitatif pur – ici de la fréquence d'occurrences des mots et des groupes de mots (ces fameux "ngrams") – comme moyen d'interprétation pour expliquer les phénomènes culturels. Et bien que je n'aie pas l'espace ici d'entrer dans toutes les insuffisances de cette méthodologie (entreprise, il faut le noter, sans aucun apport d'un humaniste), ce qu'on peut dire de positif de ce projet et de ses chercheurs, c'est qu'au moins, ils ont mis leurs données en ligne pour qu'on puisse jouer un peu avec cette collection numérique massive de livres, impensable il y a quelques années:

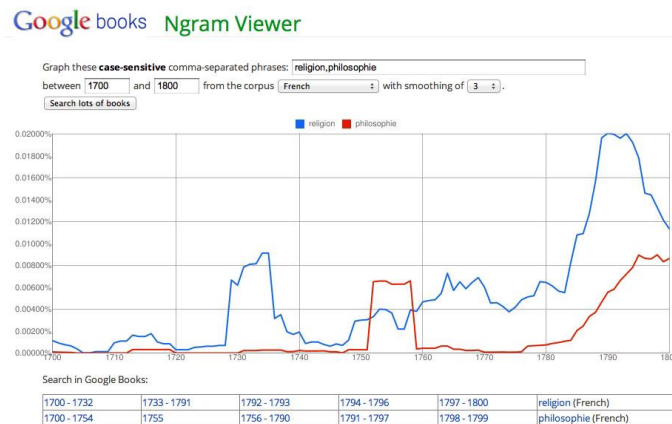


Fig. 7 : Fréquence des mots 'religion' et 'philosophie' au XVIII^{ème} siècle par Google Ngrams

Ici en fait, avec ce *Ngram Viewer* [<http://books.google.com/ngrams/>], j'ai tracée la fréquence des mots 'religion' (en bleu) et 'philosophie' (en rouge) tirée du corpus de langue française pendant le XVIII^{ème} siècle (fig. 7). Mais, qu'est-ce que ce graphique signifie ? Pour moi, en tant que spécialiste de l'*Encyclopédie*, cela me fait un certain plaisir (d'un genre tout-à-fait superficiel, je dirais) de voir ce pic de la philosophie au moment même de la publication des 7 premiers volumes de l'ouvrage monumental de Diderot et d'Alembert. Mais que dire de cette hausse des occurrences du mot 'religion' autour des années révolutionnaires ? Sans la capacité de revenir aux textes mêmes, de voir non seulement la fréquence brute des mots, mais aussi le champ lexical dans lequel ils se trouvent, il nous reste impossible de reconstruire la signification de ces

mots, arrachés, comme ils sont, à leur contexte. Il est très probable, en ce cas, que l'augmentation dans l'usage de 'philosophie' entre 1750 et 1760, aussi bien que celle de 'religion' pendant la Révolution, soient des échos des campagnes anti-philosophiques d'un côté et anticléricales de l'autre, plutôt que des virages culturels quelconques. Mais sans texte ni contexte pour situer nos mots, nous ne pouvons que rester au niveau de la conjecture.

Je ne mentionne pas ces techniques informatiques pour 'épater les littéraires' comme Baudelaire les bourgeois, mais plutôt, je les considère comme une sorte de plaidoyer pour une compréhension plus élargie, pour des dialogues interdisciplinaires plus actifs, enfin pour une collaboration plus efficace entre les sciences informatiques et les sciences humaines et sociales. Et, c'est précisément dans cet entrecroisement disciplinaire que je vois s'ouvrir un espace scientifique particulièrement riche et utile pour l'évolution des humanités numériques.

Pour ma part, j'ai eu la bonne fortune pendant ces dernières années d'être entouré à la fois d'autres 'humanistes numériques' et d'informaticiens scientifiques, que ce soit au Computation Institute et au projet ARTFL à l'Université de Chicago, ou au sein de l'e-Research Centre de l'Université d'Oxford. L'énorme avantage de cette démarche 'fédérative' de la recherche, où plusieurs équipes avec des compétences et des intérêts différents contribuent à la réalisation des mêmes projets, consiste dans l'élargissement des horizons, des méthodes, et des champs de recherches que l'on peut aborder. Cette approche fédérative à la recherche m'a donc mené à une compréhension plus profonde pour une critique plus constructive, basée sur une expérimentation prolongée et sur de nouvelles techniques informatiques dans le domaine de la fouille de données (*data mining*) et de l'apprentissage automatique (*machine learning*). Ces techniques sont empruntées par nécessité aux sciences informatiques afin de mieux explorer la bibliothèque numérique globale (ou bien infinie, pour parler comme Borges) qui est en train de se former.

Dans mes propres recherches, l'édition numérique de l'*Encyclopédie* de Diderot et d'Alembert (2013) a servi de banc d'essai idéal, en quelque sorte, pour l'expérimentation de ces nouvelles approches informatiques. L'application de ces techniques au texte encyclopédique, et les publications qui en ont résulté, nous ont aidé à mieux

comprendre la richesse de son système de classification, ainsi que la construction dialogique de son contenu (voir Horton et al. 2009, Allen et al. 2010, Horton et al. 2010). Elles offrent en sus de nouvelles possibilités d'explorer les relations complexes que les articles entretiennent entre eux ainsi qu'avec des sources extérieures. De la sorte, s'est ouvert pour moi – comme nous le verrons – tout un champ de recherche sur les citations, l'autorité textuelle, et les relations intertextuelles de l'*Encyclopédie*.

Pour ne prendre qu'un seul exemple de ce travail, et pour ne pas trop entrer dans les détails, nous avons utilisé des classificateurs automatiques (très similaires, d'ailleurs, à ceux utilisés par les filtres anti-spam des logiciels de courriel) pour explorer le système de classification de l'*Encyclopédie*. C'est-à-dire que nous avons exploité les classes existantes de l'*Encyclopédie* (à-peu-près soixante-mille articles ayant des classes assignées par des éditeurs) – où l'on voit par exemple que l'article "Beau" appartient à la classe de 'métaphysique'.

Et, avec les algorithmes classificateurs, nous avons construit un modèle de classification basé sur le vocabulaire partagé de tous les articles appartenant à une seule classe (c'est-à-dire dans le cas de 'métaphysique', tous ses 145 articles).⁴ Le modèle qui en résulte, il faut le noter, représente une réduction des 2 900 classes individuelles de l'édition originale (chiffre remarquable d'ailleurs), à un nombre-'optimal' et plus efficace pour l'algorithme de 360 classes. Ce qui soulève une fois encore cette question de la perte de spécificité inhérente aux textes humanistiques dans l'exploitation de grandes collections numériques – question que je laisse, pour le moment, ouverte.

Alors, ce modèle de classification en main, on a pu assigner une classification d'abord aux 13 000 articles qui en sont dépourvus dans l'édition originale:

⁴ Pour une explication plus détaillée de notre processus, voir Horton et al. 2009 et Horton / Roe (s.d.).

Argument, [unclassified] [Mallet] {Machine Class: Belles-Lettres}.
 Art Poétique, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Aventure, [unclassified] [Mallet] {Machine Class: Belles-Lettres}.
 Caractere d'un ouvrage,, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Charme,, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 COMIQUE, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Comique,, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 FÉERIE, [unclassified] [Cahusac] {Machine Class: Belles-Lettres}.
 MOLIERE, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Narration, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Poeme bucolique, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Poeme comique, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Poete dramatique, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Rime brisée, [unclassified] [unknown] {Machine Class: Belles-Lettres}.
 Rimes plates, [unclassified] [unknown] {Machine Class: Belles-Lettres}.

Fig. 8 : Les 15 articles non-classifiés retournés en 'Belles-Lettres'

Ci-dessus (fig. 8), nous voyons les 15 articles retournés avec la classe 'Belles-Lettres', résultats qui semblent tout-à-fait raisonnables vu les titres des articles. Et ensuite, on a aussi reclassifié, pour ainsi dire, les soixante-mille articles ayant déjà une classe (que nous avons masqué lors de cette expérience) pour interroger ce qu'on appelle l'intégrité du modèle. Et là on a vu qu'à peu près 73 pour cent des articles sont retournés avec leurs classes originelles – résultat assez étonnant, en fait, si on considère la complexité, voire l'irrégularité du système des éditeurs. Ce qui est intéressant ici, et ce qui diverge aussi le plus des approches purement informatiques (qui ne s'intéressent, pour la plupart, qu'à optimiser leurs modèles) c'est une certaine valorisation des régions d'incertitude, je dirais, soulevées par l'algorithme, c'est-à-dire les classifications modifiées, qui diffèrent et divergent de l'édition originale.

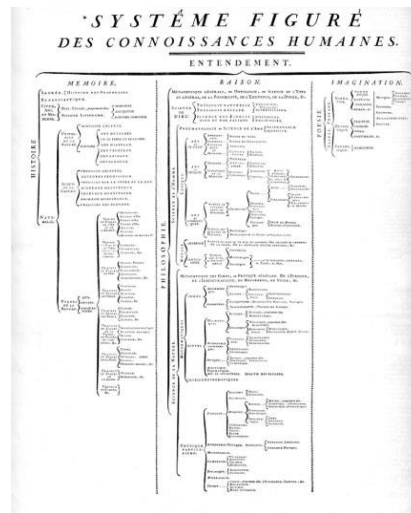
ACROSTICHE, (Belles-Lettres.) [Belles-Lettres] [unknown] {Machine Class: Poésie}.
 ACTE, (Bel. Lettres.) [Belles-Lettres] [Mallet5] {Machine Class: Grammaire}.
 Action, (Belles Lettres.) [Belles-Lettres] [Mallet5] {Machine Class: Grammaire}.
 ADAGE, (Belles-Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Jurisprudence}.
 Adjoints, (Belles-Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Grammaire}.
 AMBAGES, (Belles-Lettres.) [Belles-Lettres] [unknown] {Machine Class: Grammaire}.
 AMPHIBRAQUE, (Belles - Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Littérature}.
 ANACÉPHALÉOSE, (Belles-Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Grammaire}.
 ANACRÉONTIQUE, (Belles-Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Littérature}.
 ANAGRAMME, (Belles-Lettres.) [Belles-Lettres] [Diderot] {Machine Class: Grammaire}.
 ANTI-DACTYLE, (Belles-Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Littérature}.
 APOLOGUE, (Belles-Lett.) [Belles-Lettres] [Mallet] {Machine Class: Métaphysique}.
 APOIOPSE, (Belles-Lett.) [Belles-Lettres] [Mallet] {Machine Class: Grammaire}.
 AUTEUR, (Belles-Lett.) [Belles-Lettres] [unknown] {Machine Class: Métaphysique}.
 BUCOLIASME, (Belles-lettres.) [Belles-Lettres] [Mallet] {Machine Class: Littérature}.
 BURLESQUE, (Belles-lett.) [Belles-Lettres] [Mallet] {Machine Class: Littérature}.
 CALLIGRAPHE, (Belles-Lett.) [Belles-Lettres] [Mallet] {Machine Class: Musique}.
 CANTATE, (Belles-Lettres.) [Belles-Lettres] [Cahusac, Mallet, Rousseau] {Machine Class: Musique}.
 CANTIQUE, (Hist. & Bell. lett.) [Histoire , Belles-Lettres] [Mallet] {Machine Class: Poésie}.
 CARMEN, (Belles-Lettres.) [Belles-Lettres] [Mallet] {Machine Class: Littérature}.

Fig. 9 : Sélection d'articles en 'Belles-Lettres' avec nouvelles classes divergentes

Ici (fig. 9), en fait, nous voyons les cas où la machine a assigné aux articles originellement classés en 'Belles-Lettres' une classification différente, rendant mieux compte

peut-être du contenu d'un article ("Cantate" par exemple en Musique, ou "Acrostiche" en Poésie), ou bien, comme tout outil heuristique, incitant des réflexions plus soutenues sur un article, comme "Auteur" curieusement classé en Métaphysique, tout en soulignant la richesse que cette catégorie de 'Belles-Lettres' a eu pour les encyclopédistes.

La construction informatique de ce modèle de classification nous a fourni aussi l'occasion de comparer cette abstraction des savoirs encyclopédiques avec l'abstraction graphique du célèbre "Système figuré des connaissances humaines" que Diderot et d'Alembert ont mis en tête de leur ouvrage (fig. 10):



[Fig. 10 : Système figuré des connaissances humaines de l'Encyclopédie](#)

Comme on le sait, ce système des connaissances, censé représenter l'organisation et l'interdépendance des savoirs, a été basée sur celui de Francis Bacon, divisant l'entendement dans les facultés de mémoire (d'où l'histoire), de raison (d'où la philosophie), et d'imagination (d'où la poésie). Or, en se servant d'une autre technique tirée de l'apprentissage automatique, c'est-à-dire le groupement par des mesures de similarité lexicale (*clustering*) – et là, on peut penser aux suggestions prescientes des livres 'similaires' que l'on aimerait sans doute acheter en parcourant Amazon.com, pour avoir une idée de cette technologie. Nous avons alors regroupé les 360 classes de notre modèle par rapport à leur similarité lexicale (fig. 11):

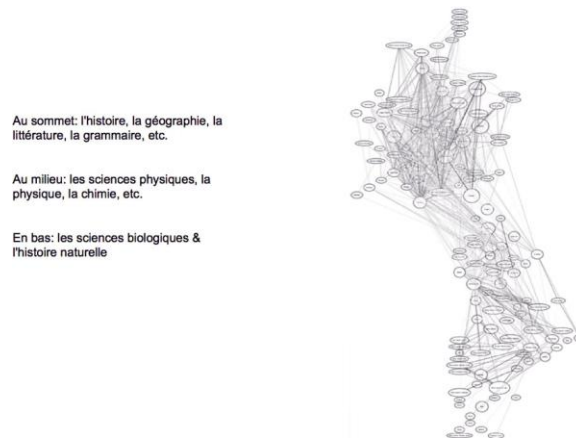


Fig. 11 : Réorganisation des disciplines de l'*Encyclopédie* par similarité lexicale

Et, à part l'illisibilité de ce graphique, ce que nous trouvons est fascinant : c'est-à-dire qu'à la différence de l'organisation facultative de Bacon et du système figuré (elle-même une sorte de vestige des systèmes médiévaux), l'organisation des savoirs que nous avons construite, à partir du texte même de l'*Encyclopédie*, ressemble beaucoup plus aux divisions disciplinaires qui se sont mises en place au XIX^{ème} siècle. C'est-à-dire, au sommet : les sciences humaines et sociales ; au milieu, les sciences physiques ; et en bas, les sciences biologiques et l'histoire naturelle. Restructuration des savoirs dans l'*Encyclopédie* qui suggère une organisation textuelle de l'ouvrage qui a été beaucoup plus tournée vers l'avenir, je dirais, que l'on ne le pensait auparavant.⁵

Toutefois, il va sans dire que ce nouveau système de classification n'est pas censé remplacer celui de Diderot et de d'Alembert, pas plus que ne le sont les nouvelles classes assignées. Le texte, dans sa forme originelle et primordiale, demeure toujours là, toujours présent à l'écran, le point de référence définitif de l'édition numérique et du lecteur. Mais, précisément à cause de son format numérique, de sa non-fixité en quelque sorte, on doit la considérer – cette édition de l'*Encyclopédie* (et toute édition numérique d'ailleurs) – comme une édition 'vivante' pour ainsi dire, toujours en train d'évoluer, de s'améliorer et de s'enrichir. Et tandis que ces compléments au texte 'quasi sacré' de Diderot choquent parfois les spécialistes, ce n'est pas, à vrai dire, pour eux

⁵ Restructuration qui va être réalisée en quelque sorte à la fin du XVIII^{ème} siècle par C.-J. Panckoucke et son *Encyclopédie méthodique*, héritière directe de celle de Diderot et d'Alembert. Voir Blanckaert / Poerret 2003, Groult 2011 et Panckoucke 2011.

que nous faisons de telles expériences.⁶ En intégrant ces techniques à l'édition numérique (au grand dam parfois de ces mêmes spécialistes), nous pensons plutôt aux étudiants, aux chercheurs, et aux lecteurs qui abordent peut-être l'*Encyclopédie* pour la première fois : en leurs mains, ces nouvelles techniques informatiques deviennent des outils heuristiques, pour ainsi dire, qui peuvent les aider à découvrir, à naviguer, et à interpréter les multiples couches dialogiques du texte encyclopédique.

En fin de compte, ces expériences à petite échelle sur l'*Encyclopédie* nous donnent alors une idée de la portée, de la promesse, et de la puissance de ces nouvelles techniques informatiques pour exploiter les collections massives de textes numériques. Mais, d'une manière plus significative peut-être, elles représentent aussi des tentatives d'exploration et de compréhension des algorithmes développés dans les sciences informatiques et appliqués à la recherche littéraire et historique. Ce genre d'expérimentation – novateur même dans le domaine des humanités numériques – sera de plus en plus important afin que les efforts de numérisation massive comme ceux de Google Books continuent à transformer nos interactions avec les archives numériques.

4 Intertexte

En forme de conclusion, j'aimerais orienter un peu cet article vers l'avenir, vers cette troisième phase dans le devenir des études littéraires et numériques toujours en train de se former : celle de la synthèse entre la textualité restreinte des collections numériques soignées et scientifiques et la contextualité grandissante des textes à l'ère de la numérisation massive. Nous entrons, si j'ose dire, dans le moment de l'intertexte; c'est-à-dire de l'interdépendance et de l'interconnexion des textes et des données.

Reflet moderne du rêve de l'interconnectivité des savoirs promu par les philosophes, cette intertextualité numérique requiert des approches synthétiques à l'étude littéraire, capables d'offrir en même temps, et la spécificité de l'œuvre d'art et la pluralité des systèmes d'information qui l'entoure ; méthodes qui facilitent, d'une manière transparente et intuitive, on peut l'espérer, le mouvement entre la macroanalyse ou la lecture à distance des collections massives et la lecture intensive ou la microanalyse des

⁶ Certains parmi ces spécialistes ont particulièrement mal compris ces nouveaux développements, s'y montrant peu disposés de saisir les enjeux méthodologiques - voire épistémologiques - du numérique et le côté expérimental de ces essais. Voir Leca-Tsiomis 2013.

textes numériques ; mobilité bilatérale et réciproque, enfin, entre le texte et le contexte par l'intermédiaire de l'intertexte.

En prenant l'intertextualité comme base théorique de nos recherches informatiques, nous nous plaçons consciemment dans la riche tradition de l'étude intertextuelle, telle qu'elle a été décrite par Julia Kristeva et Roland Barthes, ou bien de Harold Bloom, à la même époque. En suivant Barthes :

Tout texte est un *intertexte* ; d'autres textes sont présents en lui, à des niveaux variables, sous des formes plus ou moins reconnaissables : les textes de la culture antérieure et ceux de la culture environnante ; tout texte est un tissu nouveau de citations révolues. (Barthes 1968/1973: 1013)

Et ensuite Harold Bloom: "Criticism is the art of knowing the hidden roads that go from poem to poem" (Bloom 1997: 96). En s'appuyant donc sur nos expériences dans la fouille des données et l'apprentissage automatique appliqués aux grandes collections numériques, nous nous mêlons à l'investigation des approches informatiques qui identifieraient automatiquement "ce tissu de citations révolues" décrit par Barthes, dans les bases de données numériques de plus en plus massives. Et, voulant commencer avec l'aspect le plus simple, le plus concret, le plus quantifiable, de l'intertextualité numérique, nous nous sommes efforcés de repérer des passages similaires dans des œuvres différentes – approche beaucoup plus simple, il faut le dire, en théorie qu'en pratique. Pour cela nous avons employé des algorithmes qui ont été développés dans le domaine de la bio-informatique pour identifier des séquences similaires d'ADN (voir Horton / Olsen / Roe 2010).

Porté au domaine textuel, ces algorithmes ont des applications diverses, que ce soit dans la détection du plagiat, dans la collation des manuscrits ou de la critique génétique, et enfin, pour nous, dans l'identification des passages similaires entre des collections de textes.

"L'homme est né libre, et partout il est dans les fers. Tel se croit le maître des autres, qui ne laisse pas d'être plus esclave qu'eux."

trigram	sequence	bytes
homme_libre_partout	208-213	5084-31
libre_partout_fers	211-218	5098-38
partout_fers_croit	213-221	5108-46
fers_croit_maitre	218-223	5132-33
croit_maitre_laisse	221-228	5149-42
maitre_laisse_esclave	223-233	5158-58

[Fig. 12 : L'incipit du *Contrat social* de Rousseau rendu en séquences de trigrammes](#)

Dans fig. 12 nous voyons très facilement les premières phrases du *Contrat social* de Rousseau, par exemple, rendu en séquences de trois mots (un trigramme) – séquences que notre système compare automatiquement à deux textes et, dès qu'une séquence se révèle identique il enregistre le contexte, identifiant la longueur du passage avec une flexibilité de correspondance qui est, en fait, très importante vu les irrégularités des textes littéraires et l'espace informatique des collections numériques des sciences humaines et sociales.

Or, cette approche informatique nous permet d'envisager toute une gamme de recherches intertextuelles, basée sur l'identification des passages similaires dans les collections de milliers et de milliers de livres. Une fois encore, l'*Encyclopédie* nous a servi de chantier expérimental pour explorer cette technique et les enjeux intellectuels qui la sous-tendent.

En s'interrogeant sur les sources de l'*Encyclopédie* – ici, le *Dictionnaire historique et critique* de Pierre Bayle (prédécesseur avoué des philosophes) – nous avons identifié que dans l'article sur "Spinoza" [sic] attribué à l'abbé Yvon, l'auteur encyclopédique non seulement emprunte un passage de Bayle (sans citation d'ailleurs) qui réfute le monisme spinoziste, mais qu'il utilise aussi le pronom personnel de Bayle en concluant son article avec le même "Je finis par dire..." (fig. 13) :

"Spinoza, Philosophie de," Yvon? (v.15, p.474)

...ces arguments doivent convaincre la raison qu'il y a dans l'univers un autre agent que la matière qui le régit, & en dispose comme il lui plaît. C'est pourtant ce que Spinoza a entrepris de détruire. Je finis par dire que plusieurs personnes ont assuré que sa doctrine considérée même indépendamment des intérêts de la religion, a paru fort méprisable aux plus grands mathématiciens. On le croira plus facilement, si l'on se souvient de ces deux choses. l'une, qu'il n'y a point de gens qui doivent être plus persuadés de la multiplicité des substances, que ceux qui s'appliquent à la considération de l'étendue; l'autre, que la plupart de ces sçavans admettent du vuide. Or il n'y a rien de plus opposé à l'hypothèse de Spinoza, que de soutenir que tous les corps ne se touchent point, & jamais deux systèmes n'ont été plus opposés que le sien & celui des Atomistes. Il est d'accord avec Epicure en ce qui regarde la rejection de la Providence; mais dans tout le reste leurs systèmes sont comme l'eau & le feu.

Bayle, "Spinoza (Benoît de)," (T.4, p.270)

Je ne sai s'il est nécessaire que je dise que l'endroit par où j'attaque, & qui m'a paru toujours très-foible, est celui que les Spinozistes se soucient le moins de défendre. (EE) Je finis par dire que plusieurs personnes m'ont affirmé que sa Doctrine, considérée même indépendamment des intérêts de la Religion, a paru fort méprisable aux plus grands Mathématiciens de notre temps. (b) On croira cela facilement, si l'on se souvient de ces deux choses: l'une, qu'il n'y a point de gens qui doivent être plus persuadés de la multiplicité des substances que ceux qui s'appliquent à la considération de l'étendue; l'autre, que la plupart de ces Messieurs admettent du vuide. Or il n'y a rien de plus opposé à l'Hypothèse de Spinoza, que de soutenir que tous les corps ne se touchent point, & jamais deux Systèmes n'ont été plus opposés que le sien & celui des Atomistes. Il est d'accord avec Epicure en ce qui regarde la rejection de la Providence, mais dans tout le reste leurs Systèmes sont comme le feu & l'eau.

[Fig. 13 : Séquences identifiées entre l'*Encyclopédie* et le *Dictionnaire historique* de Bayle](#)

Parfois les emprunts sont plus évidents. À la fin d'un article anodin sur les "charges" de Boucher d'Argis, par exemple (fig. 14) :

Montesquieu, *De l'esprit des lois* (T.1, p.134)

Convient-il que les charges soient vénales? Elles ne doivent pas l'être dans les états despotiques, où il faut que les sujets soient placés ou déplacés dans un instant par le prince. Cette *vénalité est bonne dans les états monarchiques, parce qu'elle fait faire, comme un métier de famille, ce qu'on ne voudrait pas entreprendre pour la vertu; qu'elle destine chacun à son devoir, et rend les ordres de l'état plus permanens*. Suidas dit très bien qu'Anastase avoit fait de l'empire une espèce d'aristocratie en vendant toutes les magistratures.

"Charge," Boucher d'Argis, Diderot (v.3, p.197)

*Avant que de passer aux différens articles qui naissent de ces distinctions, nous allons exposer en peu de mots le sentiment de l'auteur de l'esprit des lois, sur la vénalité des charges [...]

L'esprit de la tyrannie est de tenir les hommes dans une oppression continuelle, afin qu'ils s'en fassent un état, & que sous ce poids leur ame perde à la longue toute énergie. 3°. Mais *cette vénalité est bonne dans les états monarchiques, parce que l'on fait comme un métier de famille ce qu'on ne feroit point par d'autres motifs; qu'elle destine chacun à son devoir; & qu'elle rend les ordres de l'état plus permanens*.

[Fig. 14 : Séquences identifiées entre *De l'esprit des lois* et l'*Encyclopédie*](#)

Diderot intervient avec son étoile éditoriale (c'est-à-dire, la marque qu'il mettait au début de tous ses articles) disant qu'avant de passer aux autres articles que ce serait une négligence de ne pas exposer ce que dit "l'auteur de l'esprit des lois sur la vénalité des charges". Citation indirecte, en quelque sorte, sans guillemets et qui ne mentionne pas le nom de Montesquieu. Mais là, nous sommes capables pour la première fois d'identifier exactement où Montesquieu commence et où Diderot finit, et aussi d'établir à partir de cette identification intertextuelle les hyperliens entre l'*Encyclopédie* et le passage exact de *L'Esprit des lois* auquel il fait référence.

Nous avons aussi pu identifier des passages dans les œuvres en traduction, ou bien tirées des bases de données non-corrigées, comme c'est le cas avec le *second traité* de John Locke tiré de la collection Gale-ECCO (fig. 15) :

Locke, John, [1783], *Du gouvernement civil* (GALE-ECCO):

Que fi le pouvoir législatif a été donné par le plus grand nombre , à une personne ou à plusieurs, teulement à vie, ou pour un tems autrement *limité; quand ce tems-là est fini, le pouvoir souverain retourne à la fociété; & quand il y ef retourné de cette manière, la fociété en peut disposer comme il lui plaît, & le remettre entre les mains de ceux qu'elle trouve bon, & ainfi établir une nouvelle forme de gouvernement* . CHAPITRE X. De l'étendue du Pouvoir législatif. IL. PAR une communauté ou un état, il ne faut donc point entendre, ni une démocratie, ni aucune autre forme pré- cife de gouvernement,t

"GOUVERNEMENT," Jaucourt (v.7, p. 789)

Si le pouvoir législatif a été donné par un peuple à une personne, ou à plusieurs à vie, ou pour un tems *limité, quand ce tems - là est fini, le pouvoir souverain retourne à la société dont il émane. Dès qu'il y est retourné, la société en peut de nouveau disposer comme il lui plaît, le remettre entre les mains de ceux qu'elle trouve bon, de la manière qu'elle juge à - propos, & ainsi ériger une nouvelle forme de gouvernement* . Que Puffendorff qualifie tant qu'il voudra toutes les sortes de gouvernemens mixtes du nom d'irréguliers , la véritable régularité sera toujours celle qui sera le plus conforme au bien des sociétés civiles.

[Fig. 15 : Séquences identifiées entre le *second traité* de Locke et l'*Encyclopédie*](#)

Ici (fig. 15), Jaucourt utilise le texte de Locke, sans citation aucune, et l'entremêle à son propos tout au long de l'article "Gouvernement".

Ces recherches préliminaires nous ont mené ainsi à une considération affinée et plus profonde sur les diverses stratégies de citation employées par les encyclopédistes. Nous avons décelé par exemple que, dans beaucoup de cas, l'acte de citer une source, ou bien de ne pas la citer, dans l'*Encyclopédie* était une fonction du statut de publication de l'œuvre citée (que ce soit publié avec privilège ou permission tacite, en anonyme ou pseudonyme, etc.). C'est-à-dire que, prenant Voltaire comme exemple typique, une citation de la *Henriade*, ouvrage inoffensif, ne valait pas celle des *Lettres philosophiques*, livre dangereux selon l'autorité, interdit et brûlé.

Et en fait (fig. 19) nous voyons par exemple, un passage tiré précisément des *Lettres philosophiques* que Jaucourt insère dans l'*Encyclopédie*, même avec attribution à Voltaire et des guillemets, mais qu'il cite néanmoins sous le titre plus acceptable de *Mélanges de littérature & de philosophie*, en évitant la mention embarrassante des *Lettres anglaises* :

Encyclopédie, "Parlement d'Angleterre," Jaucourt (v. 12, p. 141)

« Il est vrai, dit M. de Voltaire, dans ses *mélanges de littérature & de philosophie*, que c'est dans des *mers de sang que les Anglois ont noyé l'idole du pouvoir despotique; mais ils ne croient point avoir acheté trop cher leurs lois & leurs privilèges. Les autres nations n'ont pas versé moins de sang qu'eux; mais ce sang qu'elles ont répandu pour la cause de leur liberté, n'a fait que cimenter leur servitude*; une ville prend les armes pour défendre ses droits, soit en barbarie, soit en Turquie; aussi - tôt des soldats mercenaires la subjuguent, des bourreaux la punissent, & le reste du pays baise ses chaînes. Les François pensent que le gouvernement d'Angleterre est plus orageux que la mer qui l'environne, & cela est vrai; mais c'est quand le roi commence la tempête; c'est quand il veut se rendre maître du vaisseau dont il n'est que le premier pilote. Les guerres civiles de France ont été plus longues, plus cruelles, plus fécondes en crimes que celles d'Angleterre; mais de toutes ces guerres civiles, aucune n'a eu une liberté sage pour objet ».

Citation tirée des *Lettres philosophiques*, p. 90:

Il en a coûté sans doute pour établir la liberté en Angleterre; c'est dans des *mers de sang qu'on a noyé l'idole du pouvoir despotique; mais les Anglais ne croient point avoir acheté trop cher leurs lois. Les autres nations n'ont pas eu moins de troubles, n'ont pas versé moins de sang qu'eux; mais ce sang qu'elles ont répandu pour la cause de leur liberté n'a fait que cimenter leur servitude...*

[Fig. 16 : Séquence tiré des *Lettres philosophiques* de Voltaire](#)

Ces stratégies de 'non-citation' deviennent alors de plus en plus complexes, vue la légalité ou l'illégalité des œuvres qu'ils citent. Ici, Jaucourt (une fois encore) fait référence à un passage tiré du traité matérialiste *De l'esprit* d'Helvétius – livre condamné et même brûlé par le Parlement de Paris, et ouvrage, par ailleurs, indirectement responsable de la révocation du privilège de l'*Encyclopédie* elle-même en 1759 (fig. 17):

"Lacédémone, république de, "Jaucourt (v. 9, p. 153)

de législation, le plus hardi, le plus beau & le mieux lié qui ait jamais été conçu par aucun mortel. Après avoir fondu ensemble les trois pouvoirs du gouvernement, afin que l'un ne pût pas empiéter sur l'autre, il brisa tous les liens de la parenté, en déclarant tous les citoyens de Lacédémone enfans nés de l'état. C'est, dit un beau génie de ce siècle, l'unique moyen d'étouffer les vices, qu'autorise une apparence de vertu, & d'empêcher la subdivision d'un peuple en une infinité de familles ou de petites sociétés, dont les intérêts, presque toujours opposés à l'intérêt public, étendraient à la fin dans les ames toute espèce d'amour de la patrie. Pour détourner encore ce malheur, & créer une vraie république, Lycurgue mit en commun toutes les terres du pays, & les divisa en 39 mille portions égales, qu'il distribua comme à des frères républicains qui feroient leur partage.

Helvétius, Claude Adrien *De l'esprit* (p. 75)

souverain soit toujours en garde? De pareilles sollicitations, qui n'ont que trop souvent plongé les nations dans les plus grands malheurs, sont des sources intarissables de calamités: calamités auxquelles peut-être on ne peut soustraire les peuples qu'en brisant entre les hommes tous les liens de la parenté, et déclarant tous les citoyens enfans de l'état. C'est l'unique moyen d'étouffer des vices qu'autorise une apparence de vertu, d'empêcher la subdivision d'un peuple en une infinité de familles ou de petites sociétés, dont les intérêts, presque toujours opposés à l'intérêt public, étendraient à la fin dans les ames toute espèce d'amour pour la patrie. Ce que j'ai dit prouve suffisamment que, devant le tribunal d'une petite société, l'intérêt est le seul juge du mérite des actions des hommes: aussi n'ajouterois-je rien à ce que je viens de dire, si je ne m'étois

[Fig. 17 : Séquence tiré du traité *De l'esprit* d'Helvétius](#)

Sans aucune indication où la citation commence et se termine, Jaucourt insère un très subtile "comme dit un beau génie de ce siècle" pour indiquer la citation non-citée ; référence, sans doute, pour le *cognoscenti* et les autres philosophes, à l'auteur condamné.

Or, à partir des passages identifiés automatiquement entre l'*Encyclopédie* et les bases de données littéraires que nous avons consultés, nous sommes arrivés à une analyse perçante et soutenue des pratiques de citation des auteurs encyclopédiques et des sources employées lors de la composition de cette œuvre capitale (voir Roe 2013).

Enfin, ce que ce travail pointu témoigne, d'une manière efficace, ce sont les possibilités d'approches informatiques dans la recherche littéraire et historique. Ce n'est pas, en fin de compte, que les questions abordées par ces méthodes de recherche diffèrent radicalement de celles que les savants et les universitaires ont toujours posées. Au contraire, à bien des égards les questions, les matières, et les objets de recherche demeurent relativement conformes d'une époque à l'autre. Ce qui a changé dans cette transformation de la culture de l'imprimé et du livre à celle du numérique, c'est la portée des évidences, la richesse des preuves, et l'ampleur des examens que la recherche informatique apporte à l'appui de la recherche savante traditionnelle. Et si l'élaboration de ces nouvelles approches informatiques se passe actuellement dans ce domaine hétéroclite des 'humanités numériques', il est néanmoins essentiel que le développement technologique soit toujours éclairé par les questions de recherche humaniste. Ce que nous, les 'humanistes', apportons au numérique – cette profonde con-

naissance de la textualité et des enjeux linguistiques des phénomènes socio-culturels, par exemple – sera donc tout aussi important au devenir de la recherche littéraire que toute nouvelle technique issue du numérique. Et c'est précisément à partir de cette interaction vive et réciproque de la technique et de la théorie, du texte, du contexte, et de l'intertexte, que les 'humanités numériques' se définissent.

Bibliographie

- Allen, Timothy / Douard, Stéphane / Cooney, Charles / Horton, Russell / Morrissey, Robert / Olsen, Mark / Roe, Glenn / Voyer, Robert (2010): "Plundering Philosophers: Identifying Sources of the *Encyclopédie*", in: *Journal of the Association for History and Computing* 13.1. [<http://hdl.handle.net/2027/spo.3310410.0013.107>]
- Barthes, Roland (1973): "Texte (théorie du)", in: *Encyclopædia Universalis*. Vol. 15. Paris: Encyclopaedia Universalis France, 1013–17. [1968]
- Blanckaert, Claude / Porret, Michel (dir., 2003): *L'Encyclopédie méthodique (1782-1832): Des Lumières au positivisme*. Genève : Droz.
- Bloom, Harold (1997): *The Anxiety of Influence: A Theory of Poetry*. Oxford: Oxford University Press.
- Berra, Aurélien (2010): *Philologie à venir. Textes anciens et humanités numériques*. Paris: Open Edition. [<http://philologia.hypotheses.org/>]
- Chartier, Roger (1997): *Le Livre en révolution*. Paris: Textuel.
- Collectif (s.d.): "Digital Humanities Manifesto 2.0". [http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf]
- Collectif (2011): "Manifeste des Digital humanities", *THAT Camp Paris*, dir. par Marin Dacos. Marseille: Open Edition. [<http://tcp.hypotheses.org/318>]
- Compagnon, Antoine (2005): *Les Antimodernes : De Joseph de Maistre à Roland Barthes*. Paris: Gallimard.
- Crane, Gregory (2006): "What Do You Do with a Million Books?", in: *D-Lib Magazine* 12.3. [<http://www.dlib.org/dlib/march06/crane/03crane.html>]
- Diderot, Denis / d'Alembert, Jean le Rond (dir., 2013): *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, éd. par Robert Morrissey. Chicago: ARTFL Encyclopédie Project. [<http://encyclopedia.uchicago.edu/>]

- Doueih, Milad (2008): *La grande conversion numérique*. Paris: Seuil.
- Doueih, Milad (2011): *Pour un humanisme numérique*. Paris: Seuil.
- Eisenstein, Elizabeth (1979): *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press.
- Fish, Stanley (2011): "The Old Order Changeth", in: *New York Times Opinionator*, 26 déc. [<http://opinionator.blogs.nytimes.com/>]
- Fish, Stanley (2012a): "The Digital Humanities and the Transcending of Mortality", in: *New York Times Opinionator*, 9 janv. [<http://opinionator.blogs.nytimes.com/>]
- Fish, Stanley (2012b): "Mind Your P's and B's: The Digital Humanities and Interpretation", in: *New York Times Opinionator*, 23 janv. [<http://opinionator.blogs.nytimes.com/>]
- Groult, Martine (2011): *Savoir et Matières : Pensée scientifique et théorie de la connaissance de l'Encyclopédie à l'Encyclopédie méthodique*. Paris: CNRS.
- Hartog, François (2003): *Régimes d'historicité : Présentisme et expériences du temps*. Paris: Seuil.
- Horton, Russell / Morrissey, Robert / Olsen, Mark / Roe, Glenn / Voyer, Robert (2009): "Mining Eighteenth Century Ontologies: Machine Learning and Knowledge Classification in the *Encyclopédie*", in: *Digital Humanities Quarterly* 3.2. [<http://www.digitalhumanities.org/dhq/vol/3/2/000044/000044.html>]
- Horton, Russel / Olsen, Mark / Roe, Glenn (2010): "Something Borrowed: Sequence Alignment and the Identification of Similar Passages in Large Text Collections", in: *Digital Studies / Le Champ numérique* 2.1. [http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/190/235]
- Horton, Russell / Roe, Glenn (s.d.): "Machine Classifications", in: *Encyclopédie Navigational Tools*. Chicago: ARTFL Encyclopédie Project. [<https://encyclopedie.uchicago.edu/content/machine-classifications>]
- Marie Leca-Tsiomis (2013): "The Use and Abuse of the Digital Humanities in the History of Ideas: How to Study the Encyclopédie", in: *History of European Ideas* 34.4, 467–76. [DOI:10.1080/01916599.2013.774115]
- McGann, Jerome (2004): *Radiant Textuality: Literature After the World Wide Web*. New York: Palgrave MacMillan.
- Michel, Jean-Baptiste et al. (2010): "Quantitative Analysis of Culture Using Millions of Digitized Books", in: *Science* 16, 176–182 [DOI: 10.1126/science.1199644].

Moretti, Franco (2000): "Conjectures on World Literature", in: *New Left Review* 1, 54–68.

[<http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature>]

Panckoucke, Charles-Joseph (2011): *Prospectus et Mémoires de l'Encyclopédie méthodique*. Saint-Étienne: Publications de l'Université de Saint-Etienne.

Roe, Glenn / Edelstein, Dan / Morrissey, Robert (2013): "To Quote or not to Quote: Citation Strategies in the *Encyclopédie*", in: *Journal of the History of Ideas* 74.2, 213–36.

Snow, Charles Percy (1993): *The Two Cultures*. Cambridge: Cambridge University Press. [1959]

Vesser, Harold (dir., 1989): *The New Historicism*. New York: Routledge.